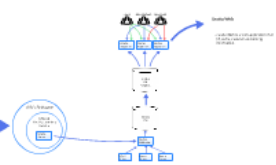




Another visualization example



Another visualization example



View of the software (OSG-UI)

Known Issues

My Contributions

Documentation

Reliability

Performance

User Interface

Openstacks Image

Google Charts

Contributors

News Release

Resources

Project Overview

Project Overview

Project Overview

Project Overview

Project Overview

Project Overview

Project Overview

Project Overview

Project Overview

Project Overview

Project Overview

Project Overview

Project Overview

Project Overview



U.S. DEPARTMENT OF
ENERGY

Open Science Grid (OSG) Gratia Web

Juan F. Mosquera
jmorales@fnal.gov

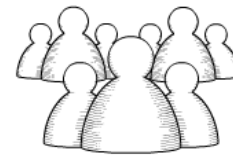


Why grid computing?

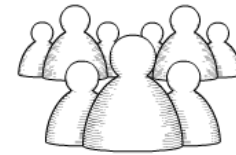
Research across sciences is not homogeneous.

Requirements of computational, and storage resources may vary even in the same field of science.

Computational needs are bursty due to the alternating of data acquisition, analysis and writing.



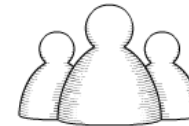
High Energy Physics



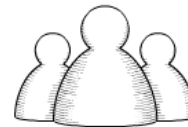
Nuclear Physics



Neuroscience



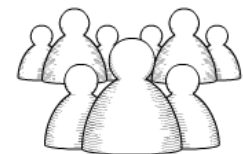
Materials Science



Astrophysics



Bioinformatics



Biological Sciences

Access to computational resources



Standalone



Limitations

Physical limitations (Ops/sec, RAM Size, etc)

Architecture limitations (x32/x64, OS, etc)

Cluster



Characteristics

Allows the distribution of jobs that can be executed in parallel over multiple machines.
Allows to have a virtual storage larger than what a single disk can offer.

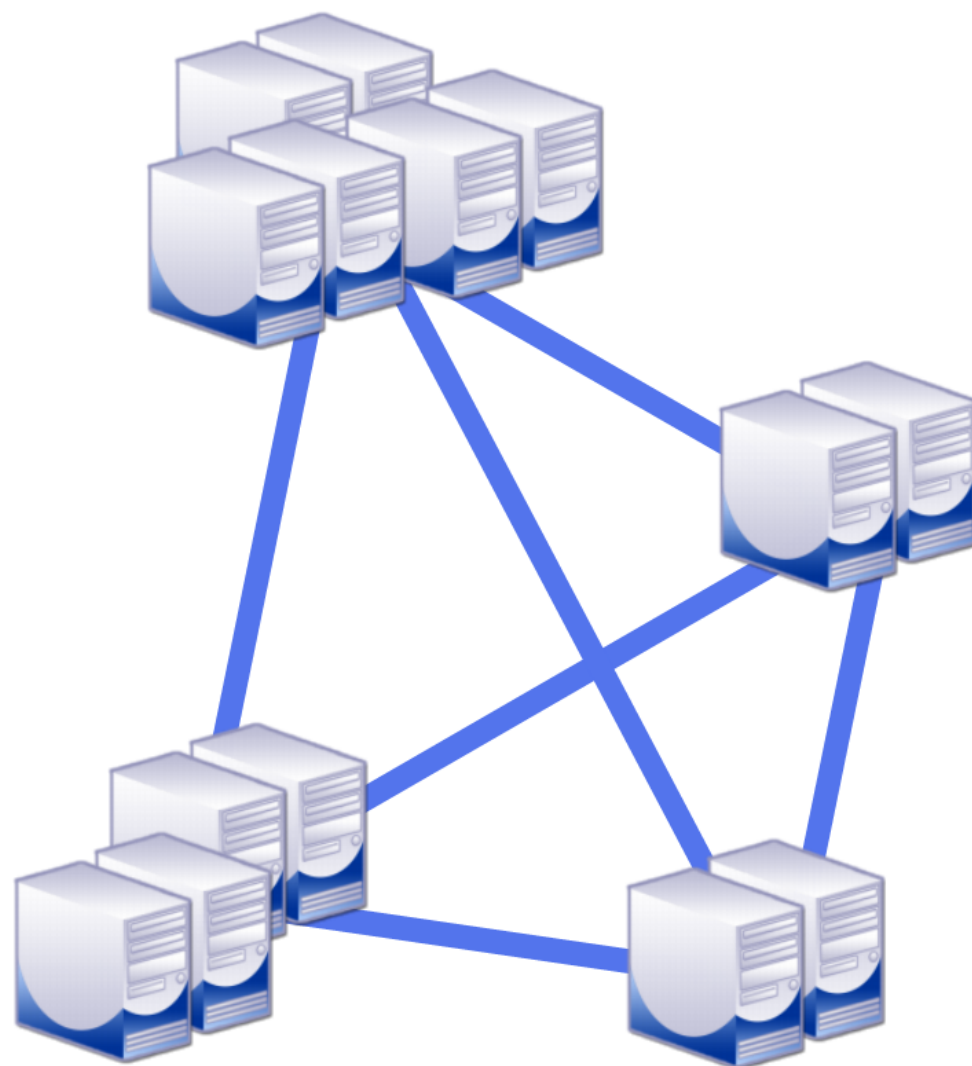
Access to computational resources



Why an Open Science Grid?

How was it organized before?

- Multiple organizations with their own cluster hardware
- Hardware dedicated to specific scientific communities or experiments
- Scientific communities or experiments were not using the hardware all the time
- Smaller scientific communities or experiments could not own/access these kind of resources



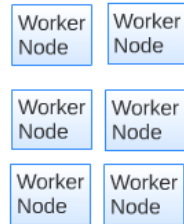
Some statistics about the OSG

- 124 Sites
- 100,000 hours of CPU per day
- 100 TB of data transfers per day



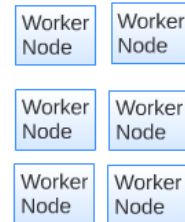
OSG

Site A



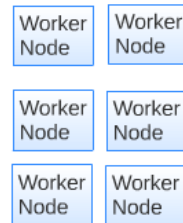
OSG
Software

Site B



OSG
Software

Site C



OSG
Software

Site B

Worker
Node

Worker
Node

Worker
Node

Worker
Node

Worker
Node

Worker
Node

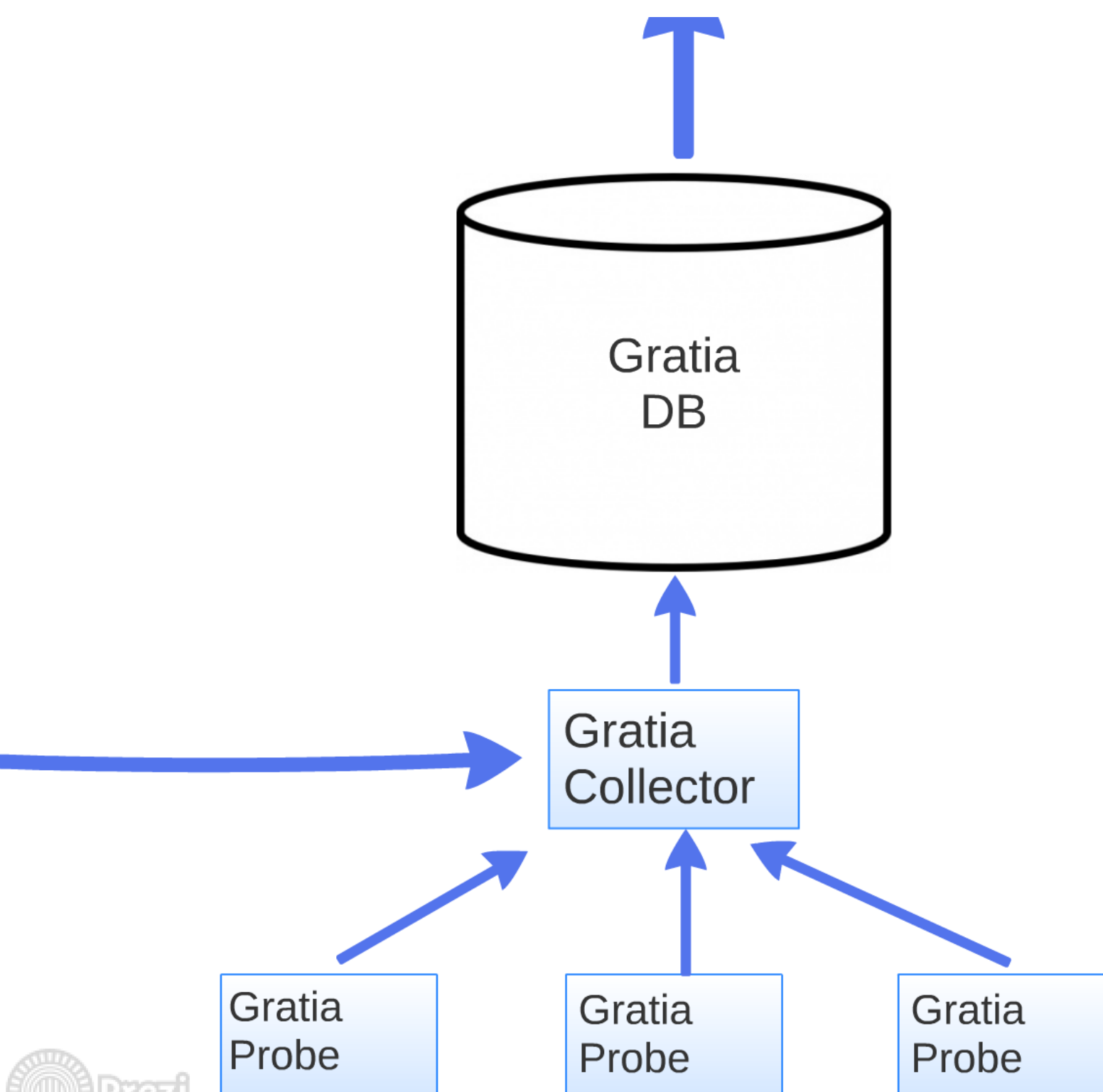
OSG
Software

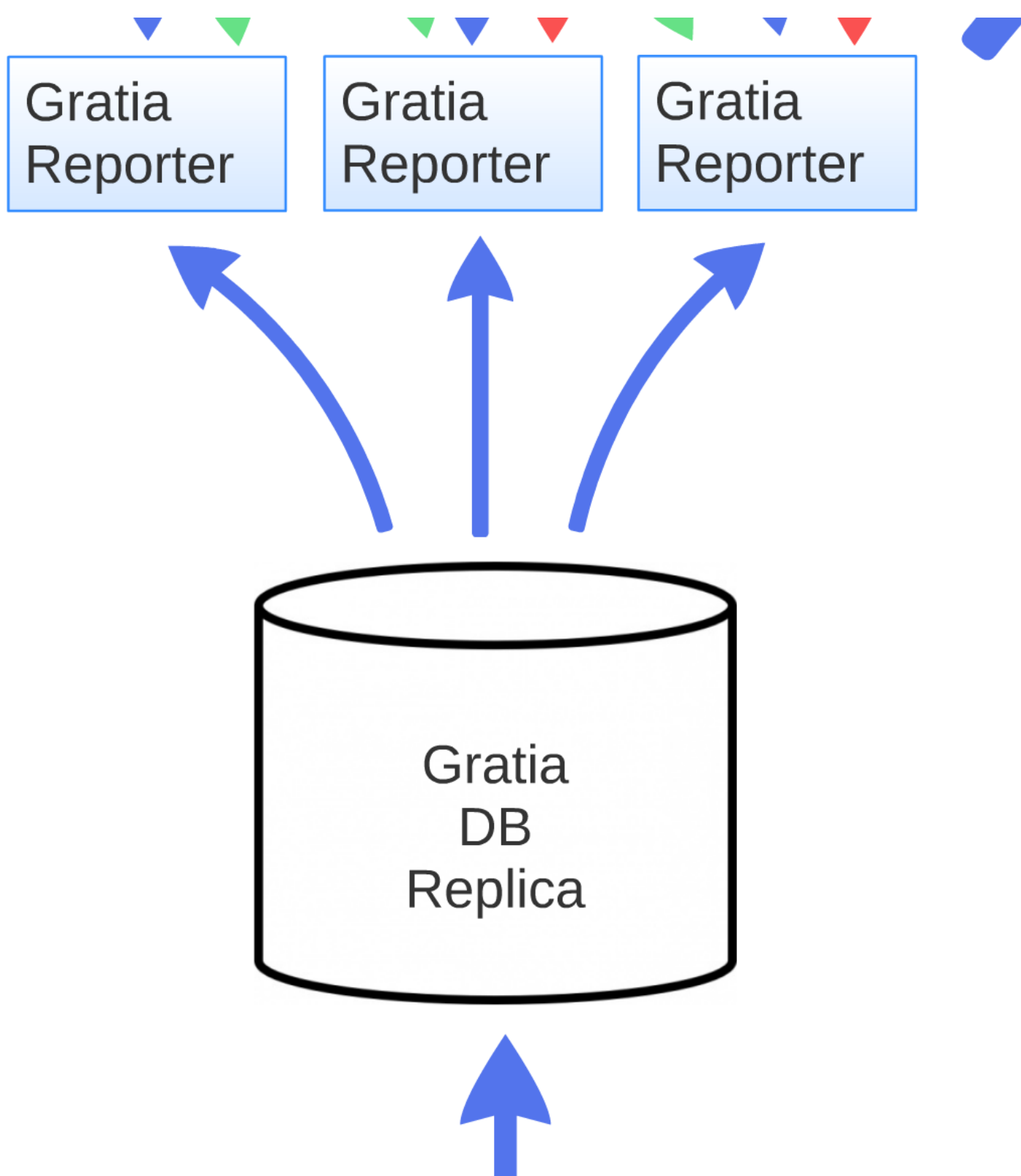


OSG Software

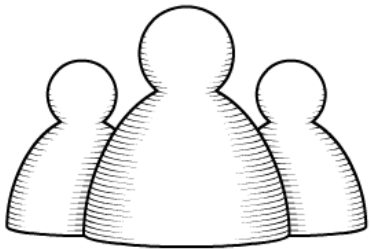
(Gratia)
Grid Accounting
Service

Gratia
Probe

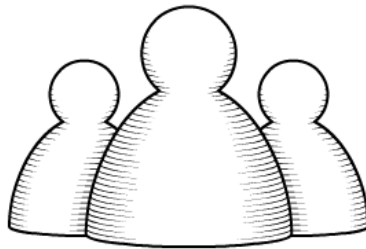




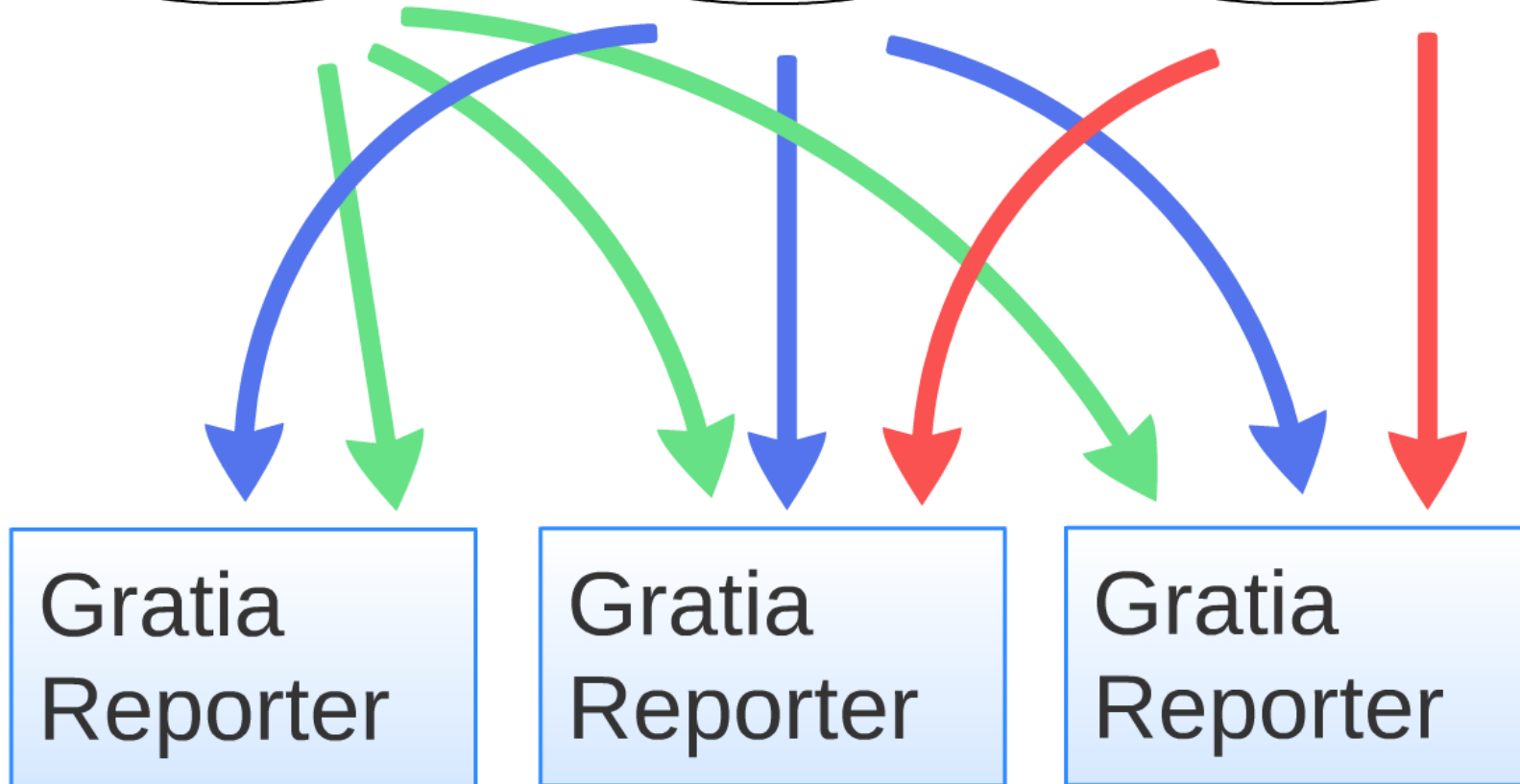
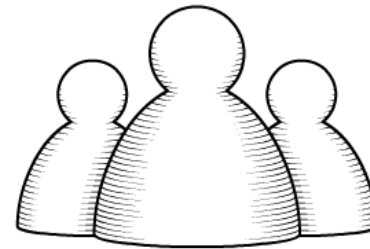
Users



Site Owners



Sponsors

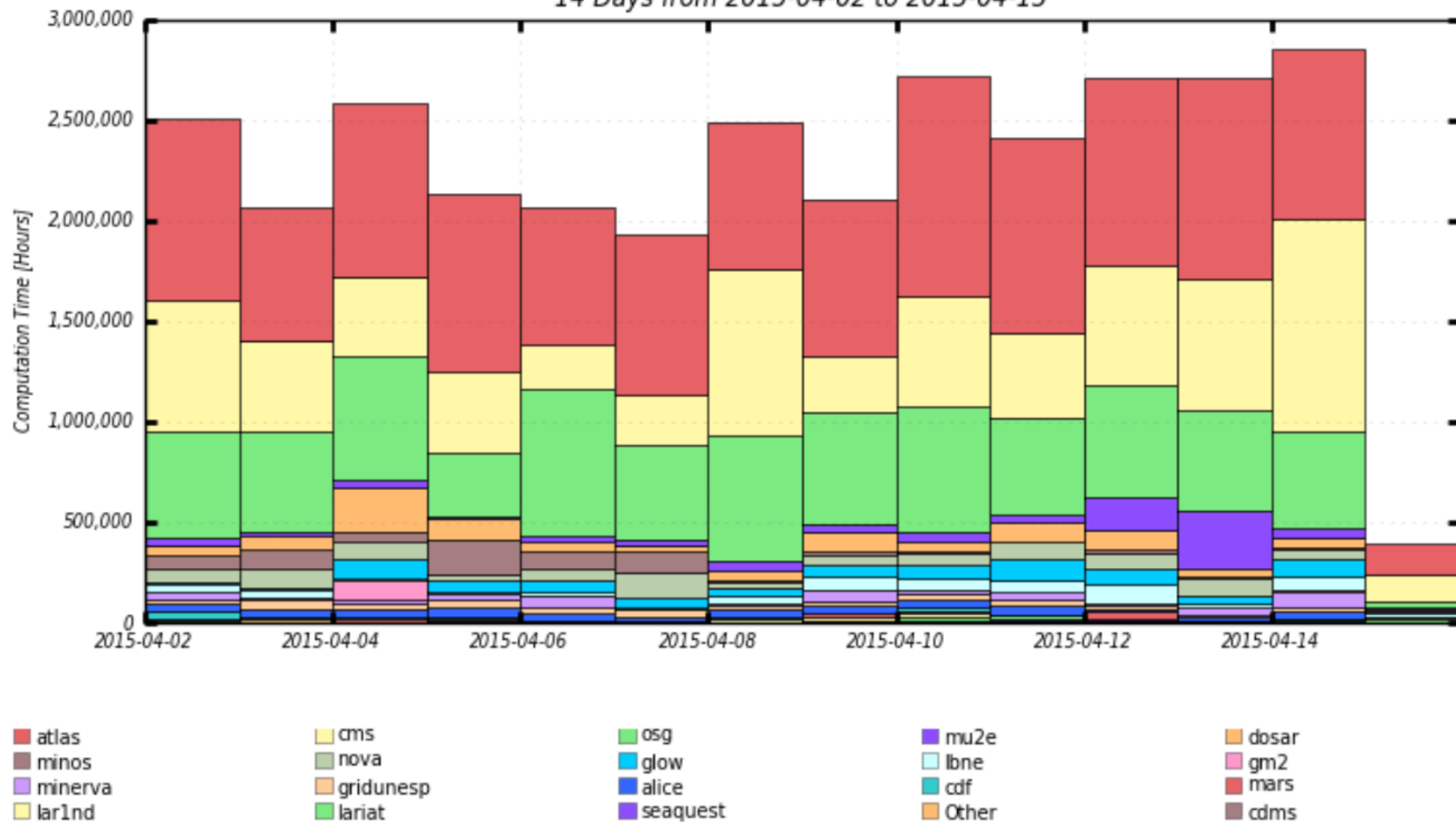


Gratia Web

- Gratia Web is a web application that let users visualize accounting information.

Gratia Web

Hours Spent on Jobs By VO
14 Days from 2015-04-02 to 2015-04-15

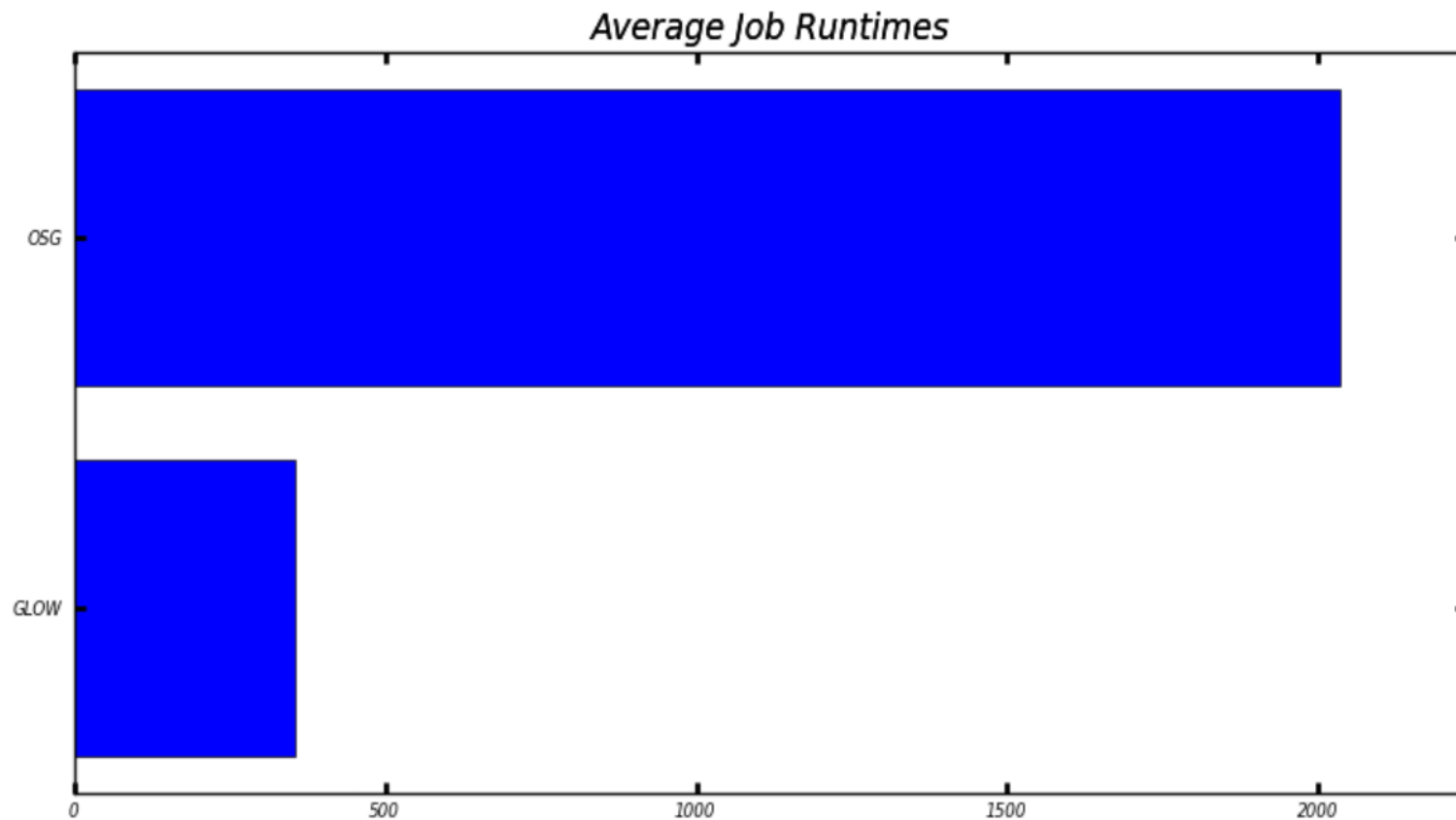


Maximum: 2,858,890 Hours, Minimum: 394,946 Hours, Average: 2,264,010 Hours, Current: 394,946 Hours

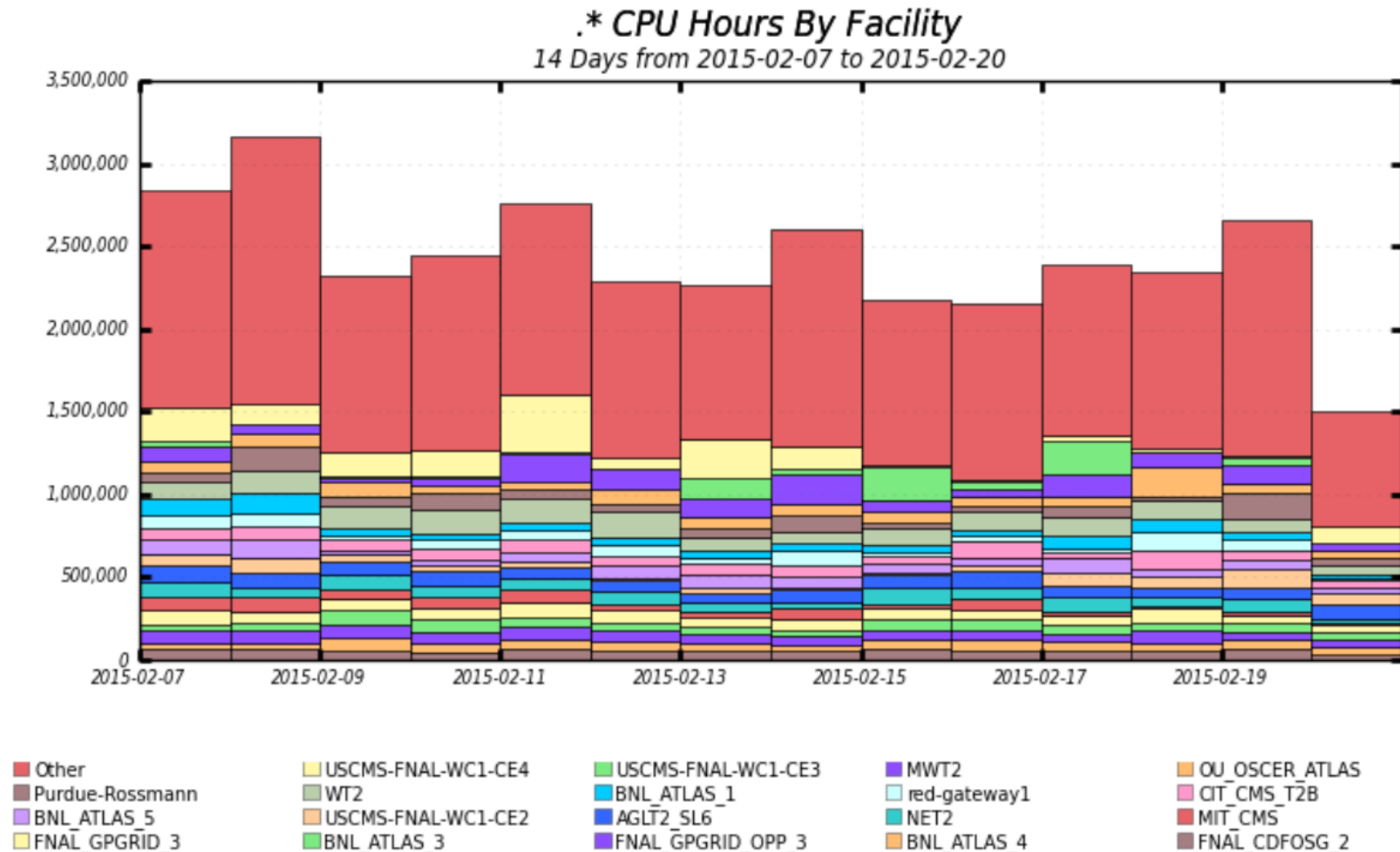
State of the software (2014-06)

- 6 Different types of plots available
 - Horizontal Bar
 - Stacked Bar
 - Stacked Line
 - Cumulative
 - Pie
 - Quality Map
- Plots are generated on the server side
- 230 Predefined views available

Horizontal Bar

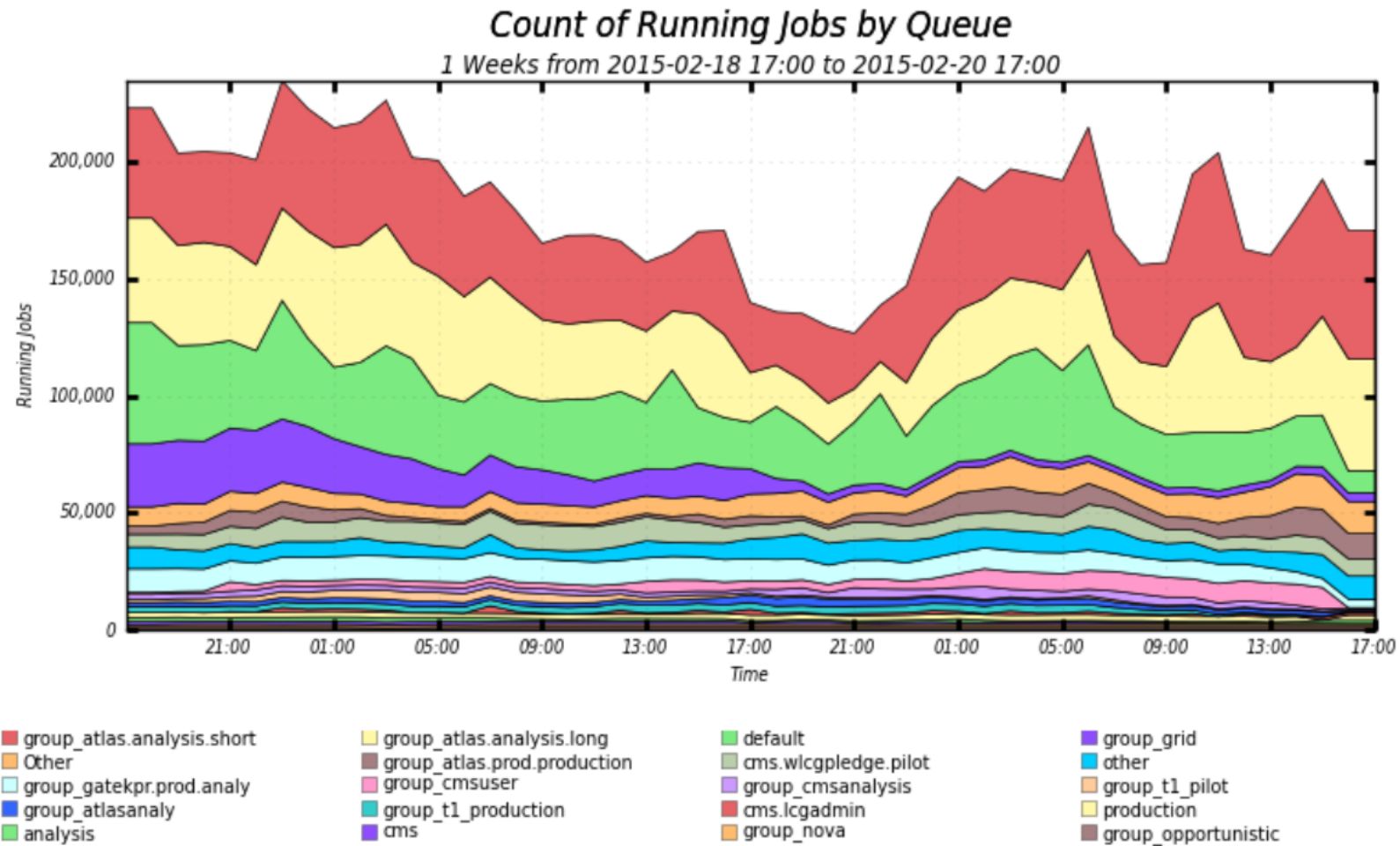


Stacked Bar



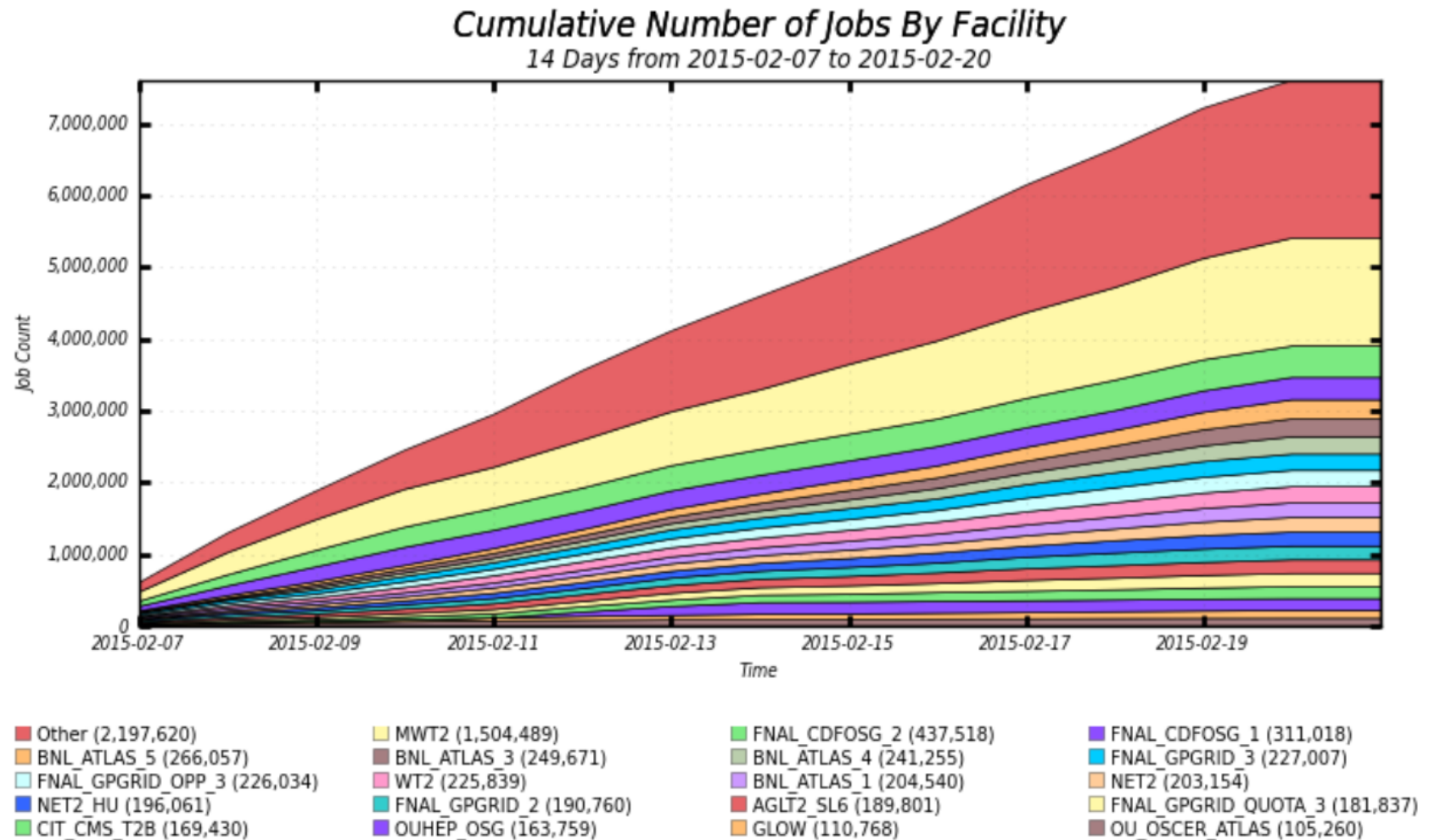
Maximum: 3,167,383 , Minimum: 1,498,245 , Average: 2,422,420 , Current: 1,498,245

Stacked Line



Maximum: 234,651 , Minimum: 0.00 , Average: 176,928 , Current: 170,679

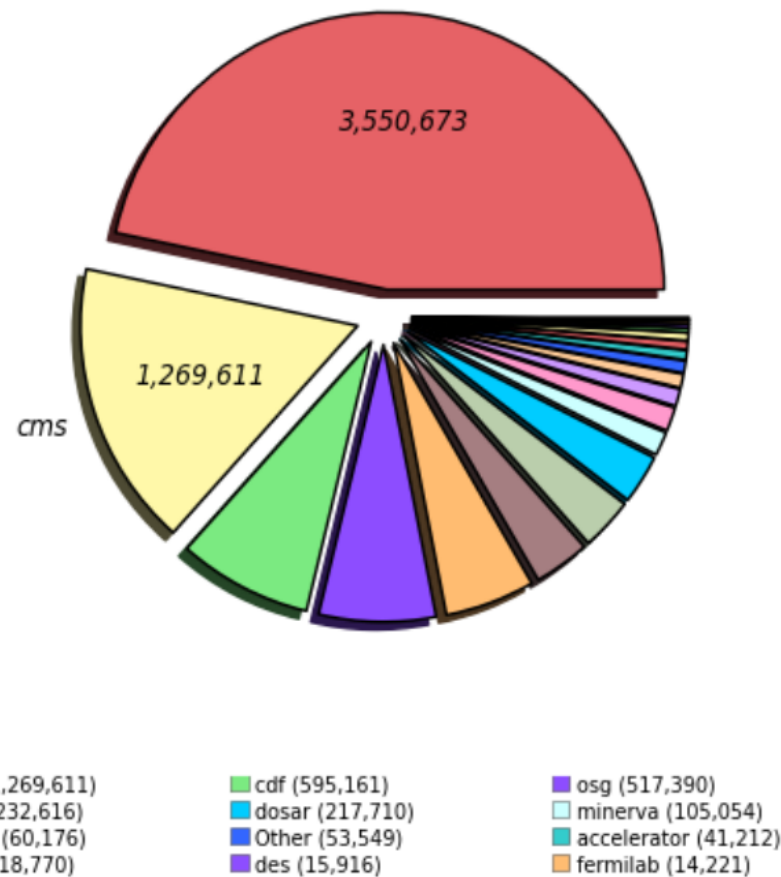
Cumulative



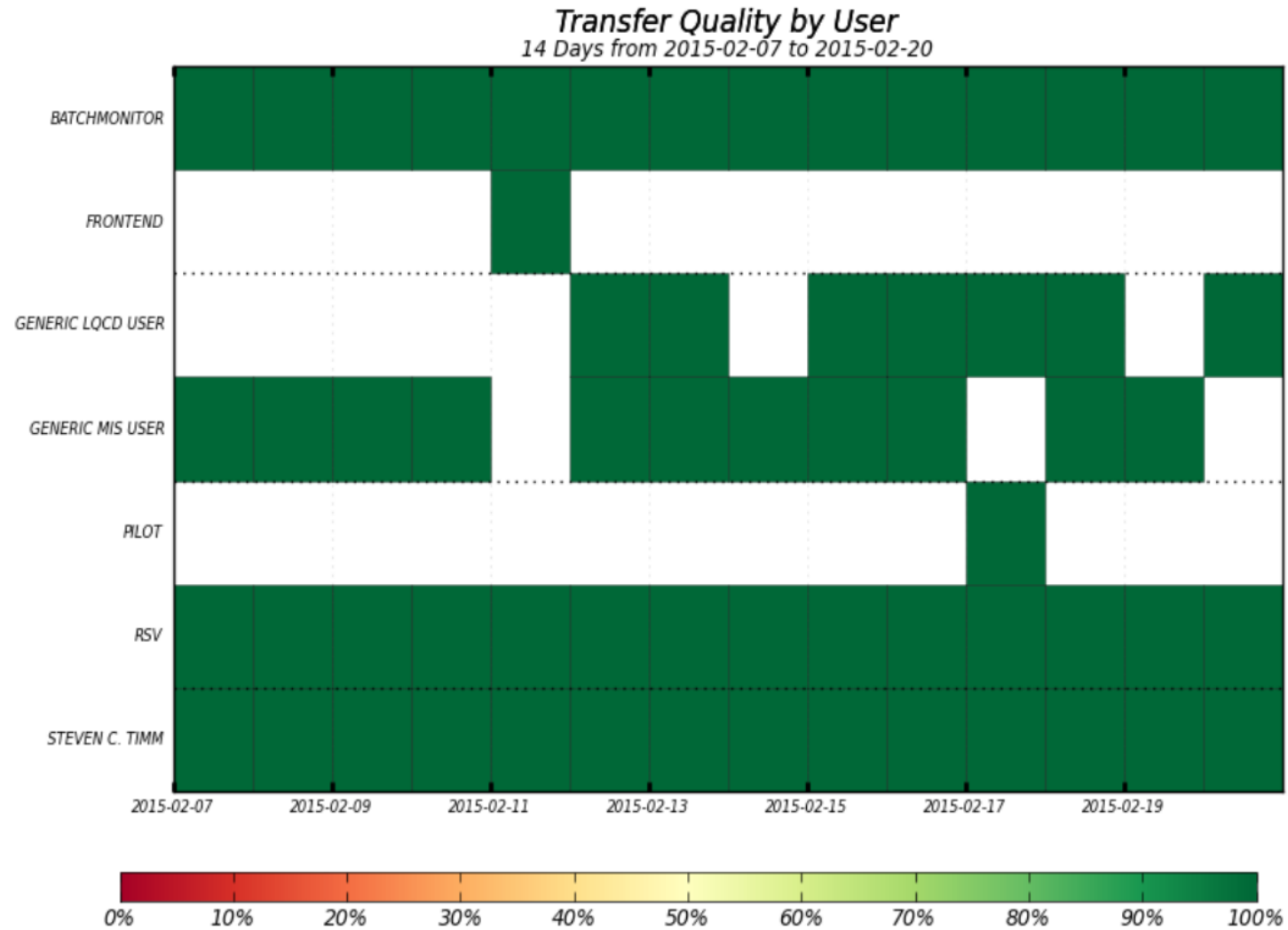
Total: 7,601,878 , Average Rate: 6.28 /s

Pie

Job Count by VO (Sum: 7,601,443)
14 Days from 2015-02-06 to 2015-02-20
atlas



Quality Map



Known Issues

- Missing documentation.
- Stability issues due to high memory usage
- Unknown issue was making the service unusable after a few days, and it required to be restarted very often.
- Software was not being maintained by OSG. However it was being used.

My Contributions

- Add documentation for users and developers
- Increase the modularity of the software package
- Improve the reliability
- Increase the performance
- Add new more dynamic visualizations(Google Charts)
- Include additional filtering/formatting in views

Documentation

- All OSG software must be documented.
However, Gratia Web didn't had documentation.
- New documentation for administrators
 - How to install and configure Gratia Web
 - How to add and customize new queries
- New documentation for developers
 - Document common practices like RPM building and preparing a release
 - Usage of Eclipse + PyDev to make development more efficient

Reliability

- Analyze resource consumption: memory, network connections.
- Fix a blocking connection bug that made the service non responsive
- Inclusion of a service in charge of prevention of connection overload of to the database.

Performance

- Some queries to the database where running slow.
- The queries where optimized to use in a better way the database indexes.
- This resulted in a three times speed up.

User Interface

- Gratia Web was made 10 years ago, and javascript support was not implemented (WEB 2.0)
- Inclusion of javascript libraries and widgets to improve user experience
 - reduce the page area by using collapsible menus
 - facilitate the selection of input values for dates and time spans
- Differentiate empty results from database connectivity, overload, and configuration errors

JQuery UI-Accordions

- [Count of Running Jobs by Facility](#)
- [OSG Current Free Space By Facility](#)
- [Count of Total Jobs by Facility](#)
- [Count of Waiting Jobs by VO](#)

Gratia Events Graphs

- [Number of Jobs Submitted by User](#)
- [Number of Jobs Started by User](#)
- [Number of Jobs Finished by User](#)
- [Number of Condor Errors by User](#)
- [Number of Events By Event Code](#)

Data Queries

- [Site Wall Clock Report](#)
- [vo_lookup](#)
- [si2k_score](#)
- [site_list](#)
- [vo_oim_corrected_table](#)
- [vo_list](#)
- [VO and Facility Usage Table](#)
- [vo_corrected_table](#)
- [vo_table](#)

▶ Gratia Events Graphs
▼ Data Queries
Site Wall Clock Report vo_lookup si2k_score site_list vo_oim_corrected_table vo_list VO and Facility Usage Table vo_corrected_table vo_table
▶ Pie Chart
▶ Glidein and Campus Grid Bar Graphs
▶ Bar Graphs
▶ Real-Time Graphs
▶ Transfer Bar Graphs
▶ Cumulative Graph
▶ GridScan Graphs

JQuery UI-Calendar

FG-IF-GRIDFTP-LAKIND.FNAL.GOV
FG-IF-GRIDFTP-MINOS
USCMS-FNAL-WC1-CE3

UCOM-OSG
rutgers-cms
FG-IF-GRIDFTP-ARGONEUT

FG-IF-GRIDFTP-
Other

Maximum: 720,111 GB, Minimum: 167,068 GB, Average: 366,926 GB, Current: 302,022 GB

Variables

Name	Value
protocol	*
facility	*
vo	*
transfer_direction(in out both)	BOTH
starttime	2015-04-14 00:00:00
exclude-facility	
span	
endtime	
exclude-vo	
Query again	

[Download results in CSV format](#)

[Show metadata](#)

[Show table of results](#)

◀

⬆

April

2015

▶

⬆

Sun	Mon	Tue	Wed	Thu	Fri	Sat	00:00
29	30	31	1	2	3	4	01:00
5	6	7	8	9	10	11	02:00
12	13	14	15	16	17	18	03:00
19	20	21	22	23	24	25	04:00
26	27	28	29	30	1	2	05:00

▼

JQuery UI-Autocomplete

Maximum: 720,111 GB, Minimum: 167,068 GB, Average: 36

Variables

Name	Value
protocol	*
facility	*
vo	*
transfer_direction(in out both)	BOTH
starttime	2015-04-14 00:00:00
exclude-facility	NONE Generic Obsolete
span	86400
endtime	3600 (1 Hour)
exclude-vo	86400 (1 Day)

Query again

[Download results in CSV format](#)
[Show metadata](#)
[Show table of results](#)

Opportunistic Usage Filtering

- In the OSG a VO can own resources where their users have priority to run jobs.
- However, users that don't own resources can use resources that are not owned by themselves if the owner is not using them. This is called opportunistic usage.
- VO Ownership of resources data is not stored inside Gratia DB.
- Gratia Web joins information from multiple databases to present relevant information.
- A new filter that lets the user easily filter Gratia Information by opportunistic and owned usage types was included.

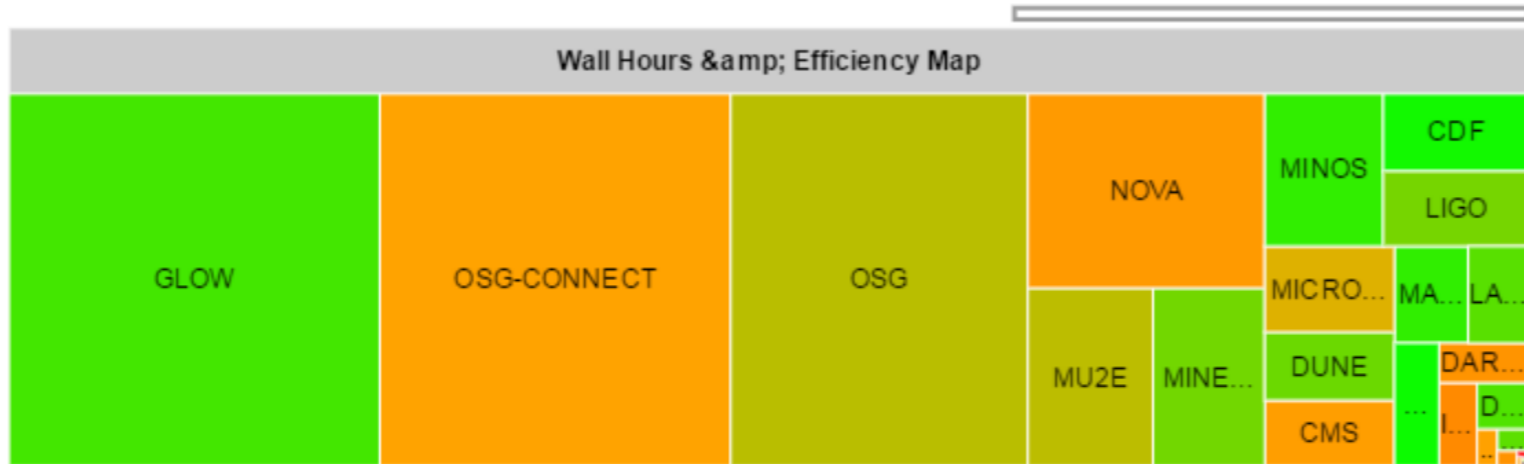
Google Charts

- Inclusion of Google Charts TreeMap
- Inclusion of Google Charts equivalents for all the visualization types available before
- Previous Matplotlib plots were not removed they are still available
- More interactive/customizable plots

Google Tree Map

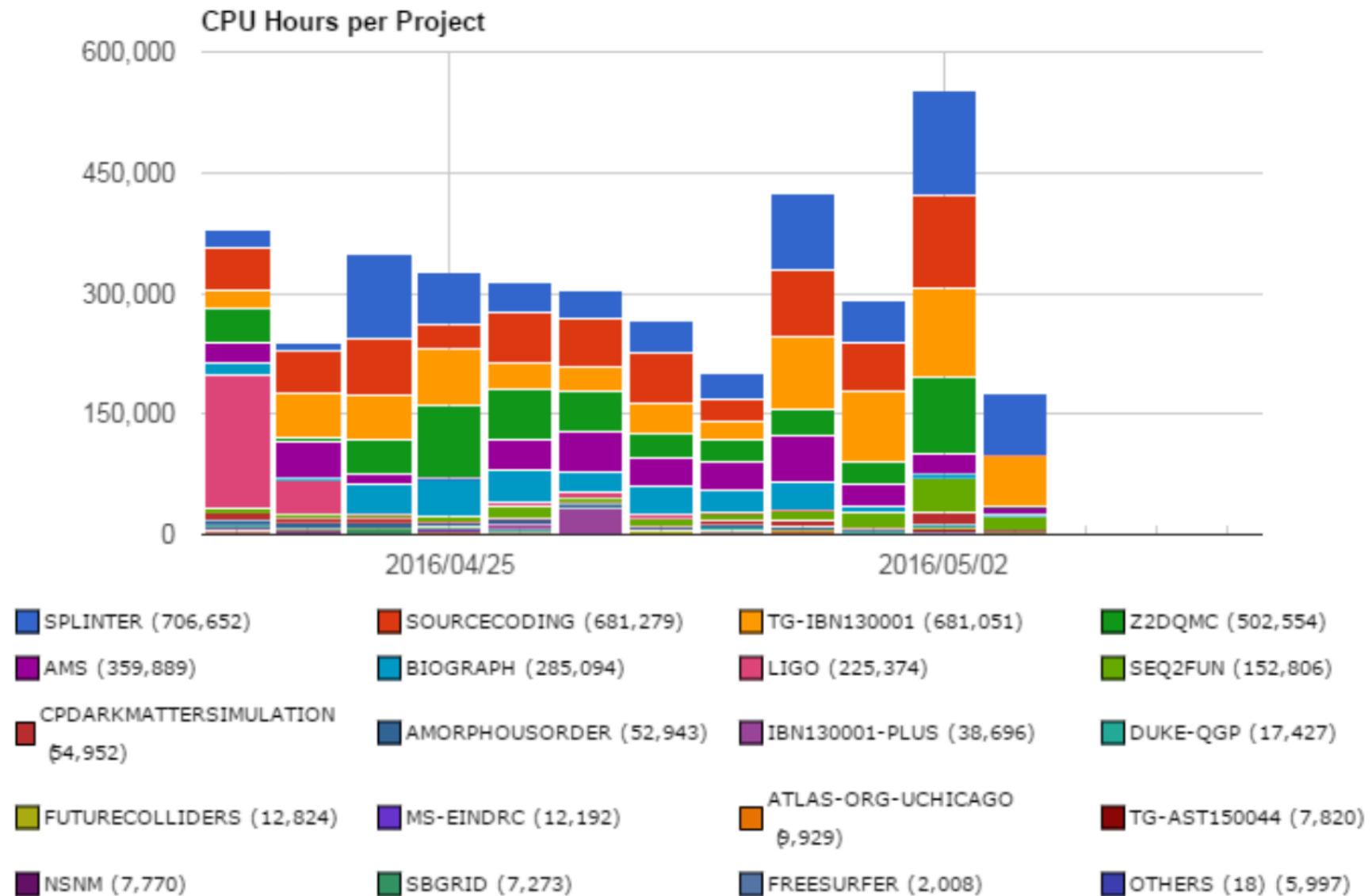
Wall Hours & Efficiency Map

Total Wall Time[Hours]: 16,010,539.10 Total Efficiency[%]: 67.45



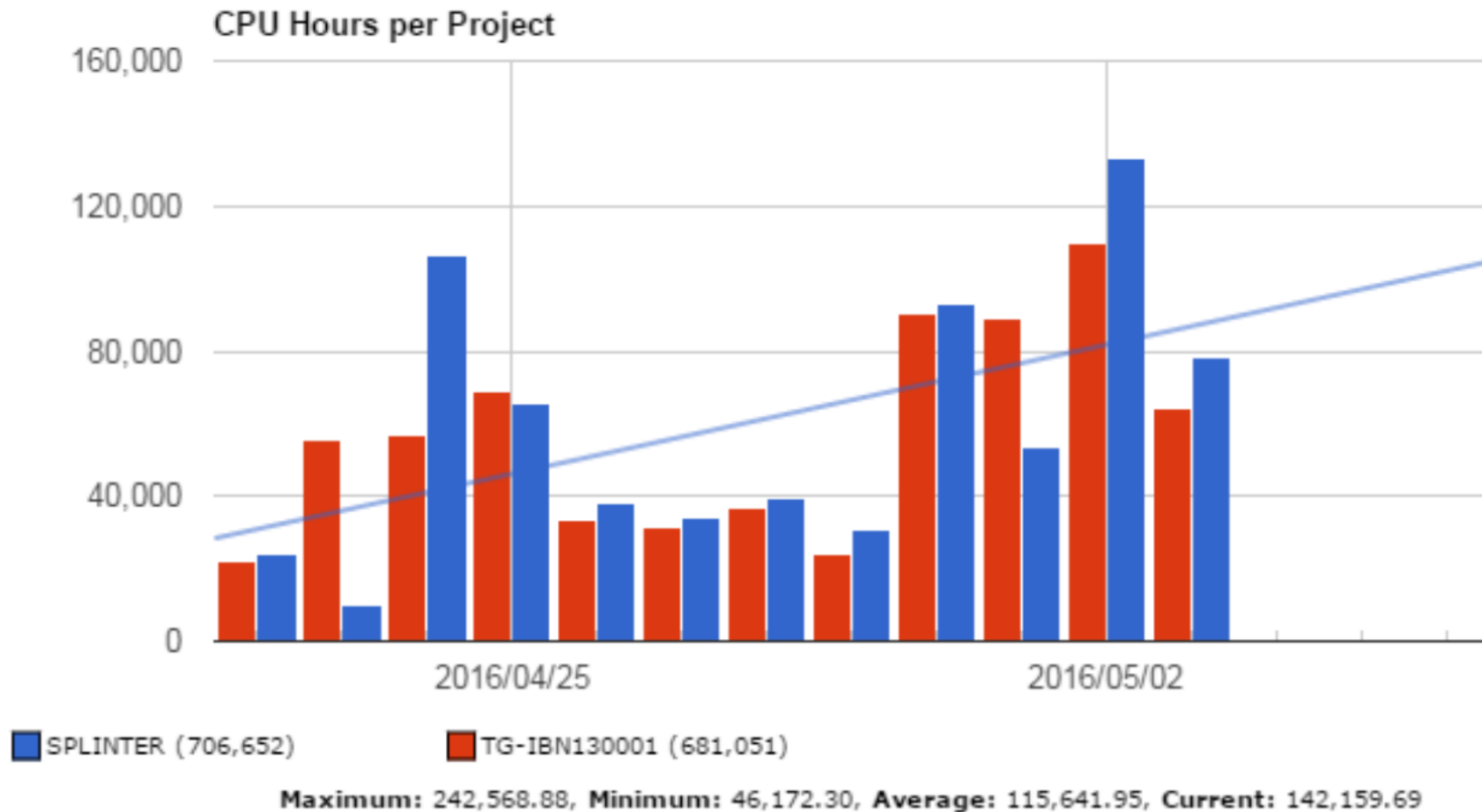
GLOW (3,900,616.03, 86.92)	OSG-CONNECT (3,696,015.12, 49.38)	OSG (3,126,995.60, 63.79)	NOVA (1,302,877.43, 46.64)
MU2E (624,366.18, 63.13)	MINERVA (568,519.89, 77.73)	MINOS (510,647.13, 90.30)	CDF (315,743.52, 96.50)
LIGO (312,652.10, 76.90)	MICROBOONE (308,827.63, 56.41)	DUNE (250,582.41, 78.95)	CMS (247,001.73, 47.86)
MARS (196,551.41, 90.64)	LAR1ND (168,185.32, 82.80)	CDMS (156,188.55, 97.60)	DARKSIDE (99,578.41, 44.83)
ICECUBE (93,381.75, 41.87)	DUKE-CONNECT (71,231.90, 84.49)	ATLAS-CONNECT (22,659.66, 48.78)	SBGRID (19,994.70, 82.02)
LARIAT (10,709.15, 44.52)	HCC (2,558.98, 0.53)	SEAQUEST (1,859.60, 47.96)	GM2 (1,221.20, 69.82)
COUPP (720.37, 15.96)	DES (604.54, 22.22)	ANNIE (137.23, 93.31)	FERMILAB (61.42, 9.00)
UCHICAGO-CONNECT (48.55, 9.88)	ATLAS (1.27, 11.70)	DZERO (0.12, 0.00)	NUMIX (0.10, 0.00)
ARGONEUT (0.09, 0.00)	BITP (0.00, 0.00)		

Stacked Bar

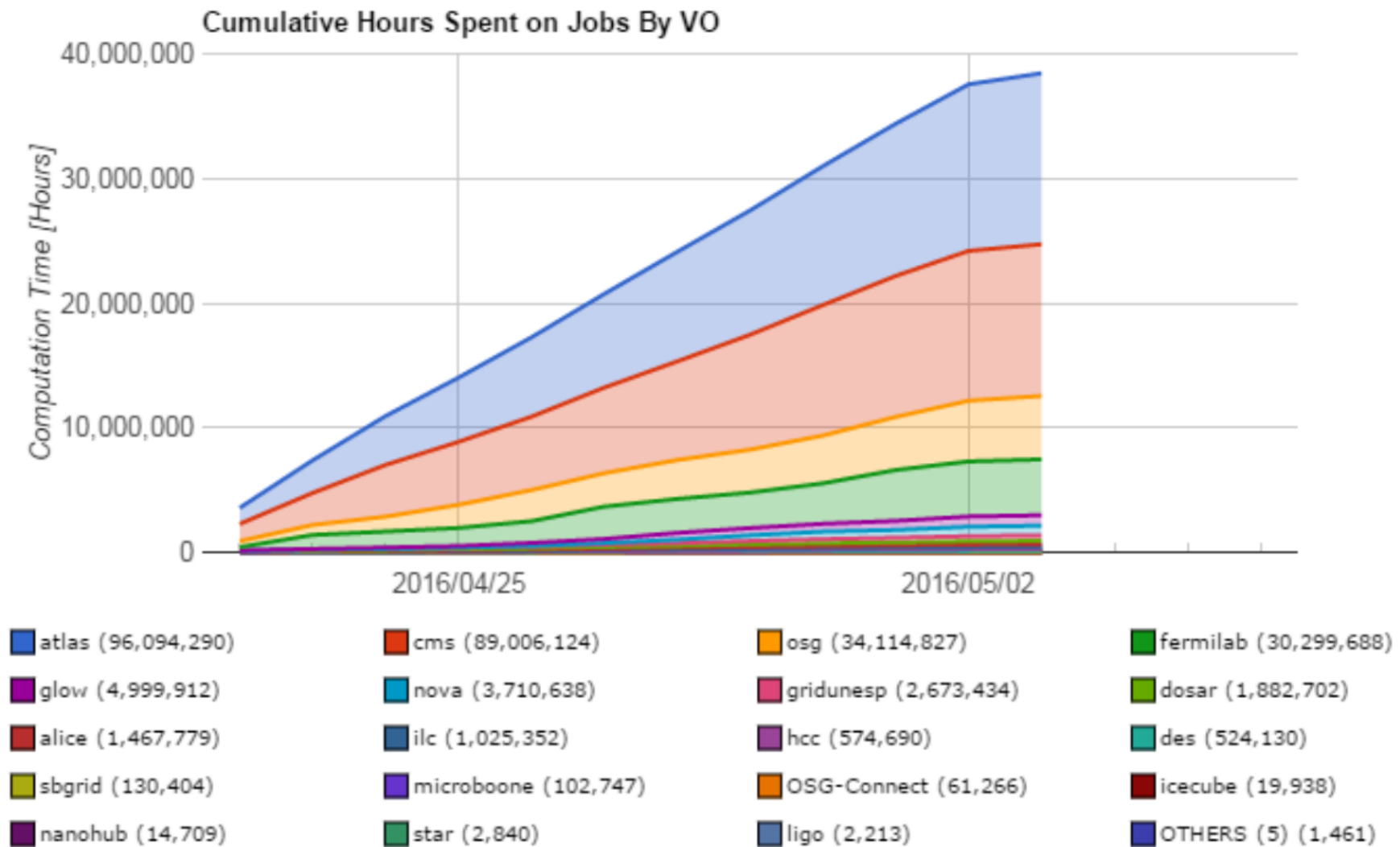


Maximum: 553,440.38, Minimum: 176,526.16, Average: 318,710.88, Current: 176,526.16

Stacked Bar Options



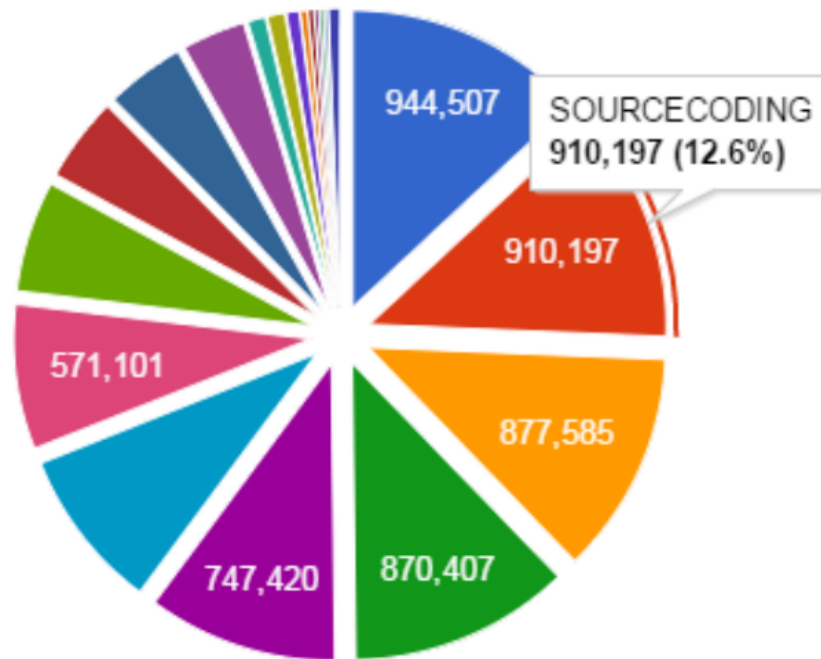
Cumulative/Stacked Line



Maximum: 38,438,643.94, Minimum: 3,561,808.54, Average: 22,225,762.01, Current: 38,438,643.94

Pie

Wall Hours by VO



TG-IBN130001 (944,507)

SOURCECODING (910,197)

SPLINTER (877,585)

Z2DQMC (870,407)

SEQ2FUN (747,420)

BIOGRAPH (638,310)

AMS (571,101)

FUTURECOLLIDERS (438,693)

TG-GEO150003 (335,981)

LIGO (312,652)

CPDARKMATTERSIMULATION
(249,615)

AMORPHOUSORDER (70,846)

DUKE-QGP (69,687)

IBN130001-PLUS (50,595)

RICEPHENOMICS (26,916)

ATLAS-ORG-UCHICAGO
(22,657)

TG-AST150044 (16,736)

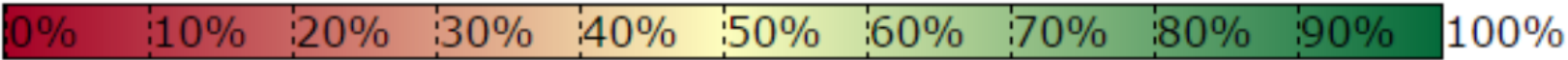
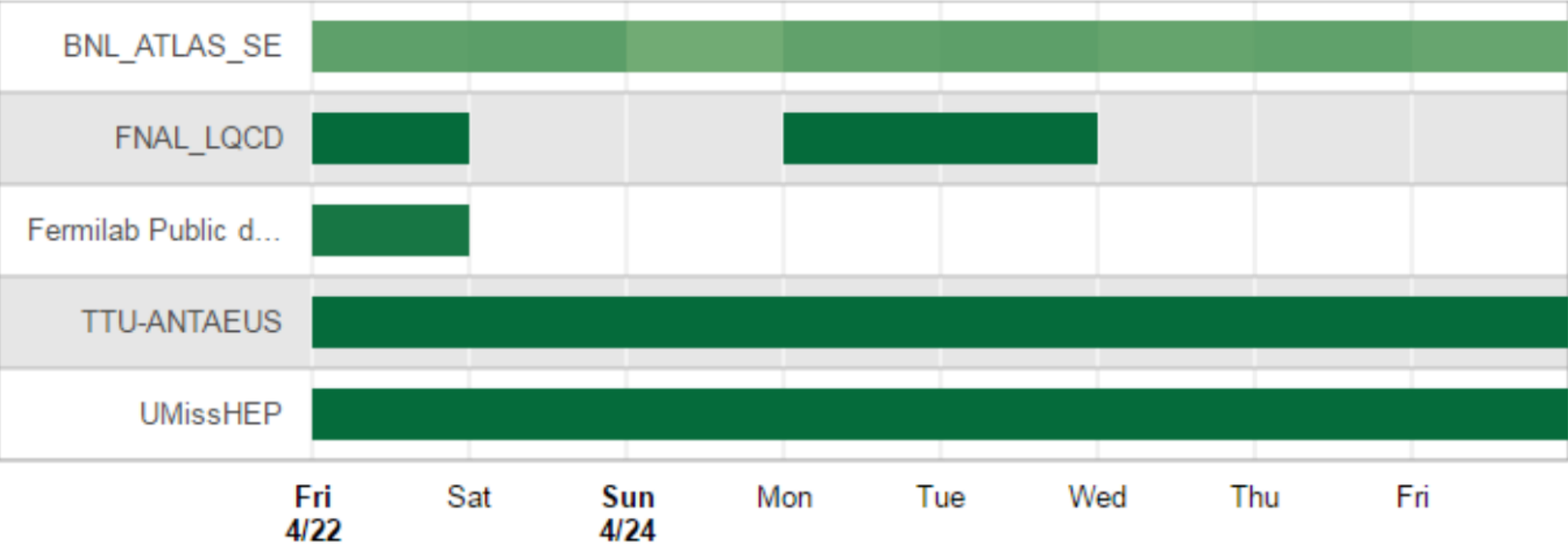
NSNM (16,687)

MS-EINDRC (15,528)

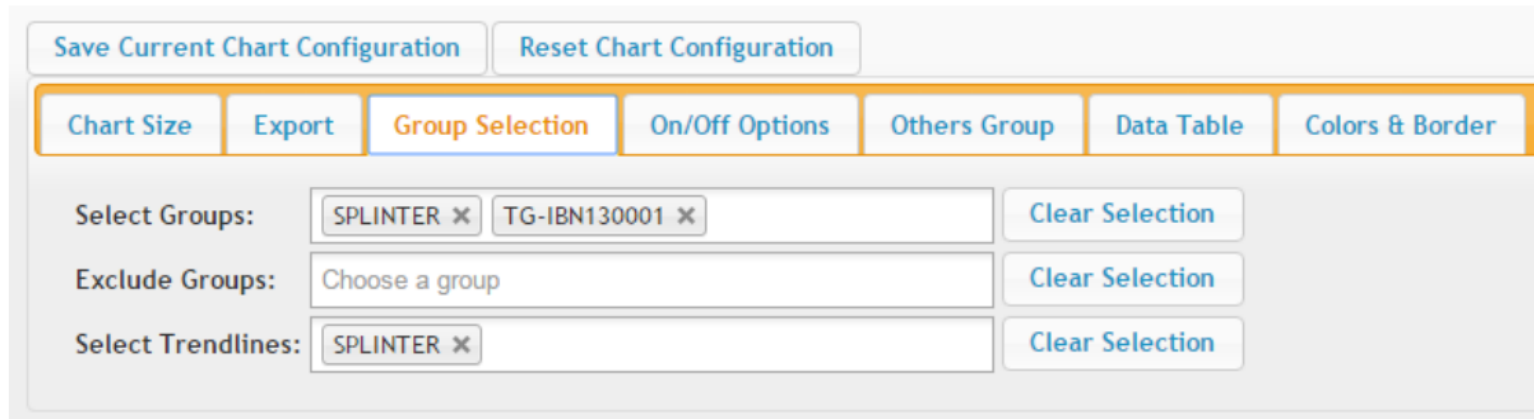
OTHERS (17) (46,033)

Quality Map

Transfer Quality by Facility



Dynamic Interaction



The screenshot shows a web application interface with a top navigation bar containing two buttons: "Save Current Chart Configuration" and "Reset Chart Configuration". Below this is a secondary navigation bar with seven tabs: "Chart Size", "Export", "Group Selection" (which is highlighted with an orange border), "On/Off Options", "Others Group", "Data Table", and "Colors & Border". The main content area of the "Group Selection" tab contains three sections: "Select Groups:" with a text input field containing "SPLINTER x" and "TG-IBN130001 x", and a "Clear Selection" button; "Exclude Groups:" with a text input field containing "Choose a group" and a "Clear Selection" button; and "Select Trendlines:" with a text input field containing "SPLINTER x" and a "Clear Selection" button.

Save Current Chart Configuration Reset Chart Configuration

Chart Size Export **Group Selection** On/Off Options Others Group Data Table Colors & Border

Select Groups: SPLINTER x TG-IBN130001 x Clear Selection

Exclude Groups: Choose a group Clear Selection

Select Trendlines: SPLINTER x Clear Selection

<http://gratiaweb-itb.grid.iu.edu/gratia/>

Conclusions

- The Gratia service collects usage information from computing resources and scientific collaborations in the Grid.
- Gratia Web allows users to visualize the collected information.
- I improved Gratia Web by:
 - Adding documentation for users and developers
 - Increasing the modularity of the software package
 - Improving the reliability
 - Increasing the performance
 - Including new more dynamic visualizations (Google Charts)
 - Including additional filtering/formatting in views

Future Directions

- Is SQL still a good tool to analyze and store the accounting data?
- Are there new technologies better suited to store/analyze/visualize the accounting Data of the OSG?

References

- [1] <http://www.opensciencegrid.org>
- [2] <http://myosg.grid.iu.edu/map>
- [3] <http://gratiaweb.grid.iu.edu/gratia/>
- [4] <http://twiki.grid.iu.edu>
- [5] [http://www.jatit.org/research/
introduction_grid_computing.htm](http://www.jatit.org/research/introduction_grid_computing.htm)
- [6] [http://cd-docdb.fnal.gov/cgi-bin/RetrieveFile?
docid=5327;filename=isgc2014Lyon.pdf](http://cd-docdb.fnal.gov/cgi-bin/RetrieveFile?docid=5327;filename=isgc2014Lyon.pdf)

Thanks

Questions?



U.S. DEPARTMENT OF
ENERGY

Juan F. Mosquera
jmorales@fnal.gov

